Merlinco

Fact Sheet No.1

MERLIN

1 Overview

MERLIN is a comprehensive software package for survey data processing. It has been developed for over forty years on a wide variety of systems, and is now one of the best-known and most widely-used packages of its kind. Its facilities include:

- data entry
- data editing and reporting
- cross tabulation with full labelling
- full and partial table manipulation
- table-based statistics
- comprehensive weighting

All these facilities are fully documented in a Windows Help File, and there is also a program entitled "MERLIN Tips and Examples", which is a library of working examples of how to use MERLIN. Merlinco offer a three-day introductory training course in the use of MERLIN which, together with the on-going support service, enables most users to analyse all but the most complex surveys. Further training courses are available for the more advanced features.

MERLIN is a "language-based" product, aimed at the professional survey analyser. In order to retain and develop skills in the use of this language, we recommend MERLIN for persons who will spend at least 75% of their time using it; for those processing data on a less frequent basis, we recommend MERLINPLUS (see below).

2 Associated products

Merlinco also market the following products, for use alongside MERLIN:

- MERLINPLUS (see fact sheet 2) – a Windows based "front-end" to MERLIN including questionnaire design and interviewing (on desktop or web). Also, a special version is available, called MERLINPLUS Database, which allows end-users of the data to do extra analysis on projects already set up in MERLIN or MERLINPLUS.

- MERLIN FASTAB (see fact sheet 3) – a *free* "data exploration" tool, simpler than MERLINPLUS Database, producing fast results with minimal specification, from projects already set up in MERLIN or MERLINPLUS.

- MORE (MERLIN On-line Reporting Engine) – a similar product to FASTAB, but *on-line*. MORE can be used to analyse data from any source, not just web interviewing.

- MERLIN TOOLKIT (see fact sheet 4) – provides a link between MERLIN and many other software packages, such as SPSS, Quantum, and SAS. The Triple-S standard is also fully supported.

- MPE (see fact sheet 5) – a complete data entry package for ASCII or card image data; it can be used for input, verification, editing, sorting, etc.

- MTWA (see booklet of examples) – a "bridge" between MERLIN(PLUS) and Microsoft Word or Adobe Acrobat, to generate high quality presentation tables that can be tailored to your own style.

3 Further information

This document summarises the main features of MERLIN.

For more information on MERLIN and other Merlinco products, contact John Tebboth at Merlinco Ltd, on +44 (0)20 7709 3000 or email info@merlinco.co.uk.

4 Operating environment

MERLIN runs under Windows Vista Service Pack and VB.NET v4.6 upwards (it may run on earlier versions of Windows but we do not support these). It interacts with Office 2007 upwards.

Simple installation instructions are provided with the software, and no "tailoring" of the operating system is needed.

5 Technical considerations

There is no practical limit on the number of questionnaires that can be handled (other than the physical size of the disk).

Data may be stored as single records up to 20 million characters long, or multiple records (such as "card images") where the sum of all the record lengths (which may vary), does not exceed 20 million. Either type may be fixed format ASCII or binary; single records can also contain comma-separated value ("csv") or Excel ("xls(x)") data, in up to 9999 fields. MERLIN uses industry standard data files which may be imported from other data entry programs, and exported into other analysis programs.

Over 30000 tables may be produced in a single run.

6 Using MERLIN

The user types instructions, in the MERLIN language, into a file using a plain ASCII text editor. The Windows "front-end" to MERLIN includes an editor, and also a spellchecker. The user may then invoke MERLIN to run some or all of the following "steps":

- "compile" to check syntax and generate internal code
- "execute" to read the data and increment tables
- "print" to generate a plain ASCII tables file ready for printing and, optionally, a "csv" or "xls(x)" file for use in Excel, or "bridge" files for exporting to Word or Acrobat (using MTWA), or E-Tabs.

Special features include the following:

- Key parameters, which vary from run to run, are presented on a screen for the user to select, to save altering the set-up file.

- House defaults may be stored in files which are automatically read on every project run.

- Labelling and formatting errors on tables may be corrected **without** re-reading the data, thus greatly reducing run times.

- MERLIN creates an "internal data file (idf)" which can be read by a subsequent set-up. This allows a project to be split into more than one set-up, so a set of tables can be produced from the "idf" or, in the case of continuous surveys, from several "idf"'s. In most cases, the "idf" occupies less disk space, and is much faster to read, than the source data.

- MERLIN also creates an "internal tables file (itf)" which can be read by a subsequent set-up. This means that additional data can be added into a set of tables without re-reading the original data. Also, using MERLIN's table manipulation facilities, it is possible to move the contents of tables so, for example, the user could add a column of data for the latest wave of data, drop a column containing the oldest wave, and move all the columns to the left.

- Each MERLIN set-up may be divided into "stages" so, for example, a single set-up could increment some tables, manipulate the aggregated figures, increment more tables weighted by figures extracted from the first set, then output a weighting factor on to the raw data file.

7 The Pre-Processor

The MERLIN compiler incorporates a program called the Pre-Processor, which is also available in standalone mode. This interprets special commands which can be embedded **anywhere** in the MERLIN set-up instructions, to save time for the user:

- Secondary files of set-up instructions may be inserted into the main set-up; these inserts may be nested.

- "DO loops" can be used to save repetitive coding. Special features can handle exceptions, such as "skip this section on the third time through the loop". Each iteration of the loop may result in different numbers, letters, Roman numerals, or **any** strings of text being substituted.

- Strings of text can be stored in data arrays, which can be retrieved within DO loops, or anywhere else; this means any string of text only needs to be typed once. The Pre-Processor can "search" for particular strings of text, or count how many strings exist; these features can be used to handle brand lists which may change. Arrays may be kept in external "csv" or "xls(x)" files.

- Data arrays may be used as a means of passing data to insert files, so they can be used like subroutines with variable arguments.

- Sections of the set-up can be skipped – conditionally or unconditionally.

- Set-ups with multiple runs (e.g. three reports, one for each area) can be handled very easily, without the need to make changes between each run, or to maintain several copies of the set-up. A single change on screen can cause the Pre-Processor to select the appropriate parts of the set-up.

8 Data input and output

- MERLIN can read **or write** data in any of the source formats, as well as its own "idf" format. It can read from two input streams of source data (e.g. to read two files in parallel), and write to 26 output streams of source data, **as well** as reading and/or writing an "idf".

- MERLIN can handle data comprising one or more physical records (e.g. card images) per respondent, some of which may be optional. The length of the physical records may vary, and MERLIN's data array may hold a total of up to 20 million columns.

- MERLIN can handle hierarchical data structures, such as household data on card 1, and each household member on card 2 which can, therefore, be duplicated any number of times.

9 Data editing and reporting facilities

- MERLIN can report errors in the data, either in a "report file", or at the screen. Another MERLIN set-up could be written to correct errors globally or individually.

- MERLIN gives the user total control over the information displayed about each error; you may display individual columns, fields, entire card images or variables, complete with messages.

- MERLIN can correct errors at run-time only, or write out a "clean" data file. It can also maintain "clean" and "dirty" files, so only "dirty" data is submitted to error correcting and re-checking runs.

- MERLIN can produce top-line counts of source data, both of individual codes (hole counts) or fields containing numeric or character data (frequency counts).

- MERLIN has extensive facilities for listing disaggregated data. Many formatting options make it possible to produce presentation quality documents in which, if desired, codes can be replaced by labels, e.g. "Respondent 123: Sex is male, Class is AB". It is thus possible to produced personalised letters or forms, including values from the data file.

- MERLIN can produce "report files" that are capable of being read by another program, i.e. without page breaks or headers, in fixed ASCII, "csv" or "xls(x)" formats.

10 Variable definitions

- MERLIN supports five types of variable:
    - single response (up to 32767 bits with 0-1 bits true)
    - multiple response (ditto with 0-4096 bits true)
    - numeric without decimal point (plus or minus 2,000,000,000)
    - numeric with decimal point (plus or minus 13 significant digits)
    - character (up to 4095 characters, extended character set allowed)

- Additionally, the following may be viewed as special types of variable. Most of the operations allowed with the five main types are also allowed with these, except that they may **not** be tabulated (they must first be converted into one of the five types):
    - a column (a single column viewed as a multiple response variable)
    - a field (consecutive columns viewed as numeric or character data)
    - a spreadfield (consecutive fields used to store multiple response data)
    - a freefield (a field in a "csv" or "xls(x)" file)

- Variables may be re-defined as often as required, providing type and length are not changed. Individual bits of single and multiple response variables may be set on or off.

- "Net" bits, to summarise data, and "reject" bits, can easily be created.

- Many mathematical functions are available, such as random number generation, square roots, plus functions for doing date and time calculations.

- Up to 124 external "lookup" files in "csv" or "xls(x)" format may be used to check a value in the data, and/or to extract other values related to it.

- Variables may be given different sets of labels, to be used when tabulated vertically, horizontally, or used in a filter.

11 Branching facilities

MERLIN offers the user many facilities for "branching", i.e. filtering selected variable definitions and tables on some records only. These facilities include:

- Simple IF statements (applying to one statement only)

- IF blocks (applying to a set of statements), which may contain "elseif" and "else" clauses. IF blocks may be nested.

- UNLESS statements and blocks, as an alternative to IF – sometimes useful for complex negative conditions.

- FILTER statements – similar to IF, except they automatically pick up a label associated with the filter condition specified.

- LOOP statements for generating run-time "DO loops". These may be nested.

- BREAK statements for exiting from loops prematurely.

- GO TO and "computed GO TO" statements, for branching to specific places in the set-up depending on a logical condition, or the value of a variable. GO TO statements may branch backwards.

- GO OUT TO and GO BACK statements for creating "in-line" subroutines. This allows the user to branch to the same section of code from anywhere in the set-up, then return to the point from which the branch was made.

12 Tables

- MERLIN can produce printable tables with up to 32000 rows and 1500 columns. It is also possible to "join" printable tables vertically.

- MERLIN tables can store up to thirteen significant digits of precision.

- The vertical axis may contain distribution data from single or multiple response, character or numeric variables, and statistics calculated from these. The horizontal axis may contain distribution data from single or multiple response variables, and statistics calculated from these.

- Summary tables may be constructed, a row at a time, a column at a time, or even a cell at a time, and tables of statistics only may be produced.

- "Net rows" can be created automatically, replacing all rows where the data are below a specified percentage of the total.

- Table cells may be incremented more than once to add data together, e.g. to produce a table based on "all items" rather than "all respondents".

- Tables may be weighted by factors supplied by the user, factors in the data itself, or factors calculated by MERLIN to achieve target figures supplied. MERLIN supports "multi-stage" and "rim" weighting.

- Each cell of a MERLIN table may contain any combination of unweighted and weighted numbers, column, row and total percentages, and indices – shown underneath or alongside each other.

- Blank rows and columns can be suppressed, and rows may be ranked in ascending or descending order, using the values in any of the columns. Multi-level ranking is allowed to show, for example, ranked models within ranked items within ranked manufacturers.

- There are about 300 options that control the appearance of the tables – how tables and pages are numbered, which titles are printed and where, which characters are used for empty and nearly empty cells, how many decimal places are printed, to name but a few.

- Formatting options may be applied globally, to individual tables, or to parts of a table so, for example, one column may be wider than the rest, or have a different number of decimal places shown.

- Statistical functions include means, standard deviation, error variance, standard error, mode, median and other quantiles, chi-square, independent and dependent t-tests, correlations, f-tests, significance tests, Kolmogorov-Smirnov and Mann-Whitney-Wilcoxon tests.

- MERLIN tables will use labels assigned to the variables being tabulated, or directly to the tables. These labels include headers and footers, table and page numbers, vertical, horizontal and base titles, and row and column labels. MERLIN automatically splits and formats labels into separate lines where necessary, or the user may control this himself.

- MERLIN automatically splits tables that overflow the page vertically or horizontally, repeating headings, and generating continuation numbers, etc. It will ensure that tables are split between defined "groups" of rows or columns, or the user may specify the break points himself.

- MERLIN automatically produces a table of contents describing the tables produced.

13 Table manipulation

MERLIN has extensive facilities for manipulating tables after incrementation, similar to those offered by spreadsheet packages; these include the following:

- MERLIN can add, subtract, multiply or divide, entire tables, rows, columns, cells, or any block of cells within the same or different tables.

- MERLIN table manipulation can access not only the raw aggregated data, but any statistics specified.

- Text may be inserted into table cells, e.g. "high" or "low", depending on the value found in the cell.

- Constants may be inserted into table cells, and values **from** table cells may be extracted into variables.

- Several functions are available to facilitate table manipulation, e.g. to sum all the values in a table, row, column, or block of cells, or to replace values that fall between specified limits.

- Additional functions make it possible to manipulate multiple tables on a global basis, without specific reference to the table names.

- Tables may be used to store "look-up" data, and a function is provided to easily extract such data.

14 Exporting tables

MERLIN tables can be output as comma-separated ("csv") or Excel ("xls(x)") files, which can be read directly into Microsoft Excel. They can also be output in a TabsML ("xml") file for exporting into E-Tabs, or in an "mtw" file for exporting into Microsoft Word or Adobe Acrobat via MTWA (both available from Merlinco).